Cinematic AI:

Find The Movies You'll Love

Explore the captivating world of cinema through a targeted browsing experience powered by:

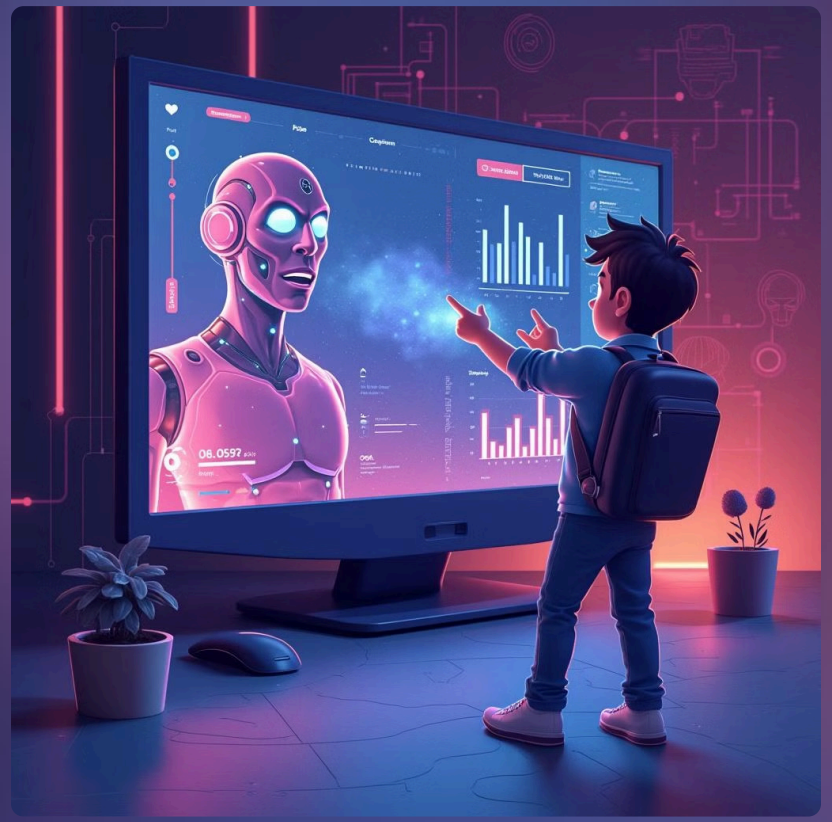
- Sentiment Analysis
- Emotion Detection
- Similarity Recommendations

Made with Gamma

Product Overview

Our application creates an all-in-one interface that allows customers to access both movies insights and recommendations on the same platform. Here is an overview of the process:

1. User is prompted to select a movie
2. System locates movie reviews based on selection and analyzes the reviews based on sentiment
3. Emotion Detection Algorithm classifies the description of the movie based on emotion
4. Recommendation system provides with a cinematic body of work based on movie selection



Made with Gamma

Model Setup and Training

Data Collection & Preprocessing

- DataFrame: comprised on Sentences and corresponding Emotion
- Preprocessing & Label Encoding
  - Remove whitespace and splitting sentences based on delimiter
  - Emotion Labels encoded into format suitable for training.

Training & Testing

- **Train-validation split technique** used for effective learning and validation of performance
- Callback techniques used to monitor training and prevent overfitting
  - **ModelCheckpoint**
  - **EarlyStopping**

Interpretation & Results

- Sentiment Analysis
- Emotion Detection
  - Joy
  - Sadness
  - Anger
  - Fear
  - Love
  - Surprise

Made with Gamma

High-Dimension Sentiment Analysis

1 Support Vector Machine (SVM)

- Vectorization
  - TF-IDF
- Kernel Function
  - Linear
- Hyperplane Classification

2 Feature Extraction

- Convert text data into numerical form
- TF-IDF Vectorization
  - Frequency of term in dataset

3 SVM Sentiment

- **Hyperplanes** (decision boundaries) in high-dimensional space
- Uses the **largest margins** to classify the positive and negative data points

4 Kernel Trick

- **Mathematical function**
- Employed to reorganize dataset for **nonlinear classification**

5 Required Expertise

- Parameter Tuning
  - Kernel Selection
  - Regularization Parameters

6 Challenges

- Interpretability of SVM Models with non-linear kernels
- Decisions Trees provide better BI visuals

Made with Gamma

The Layers of Emotion Detection

1 Deep Neural Architecture

- Input & Attention Masks
- BERT Embedding
- Global Average Pooling1D
- Dense Layer
- Dropout Layer
- Output Dense Layer

2 Inputs & Attention Mask

- Input layer structured to take in sequence of up to 100 tokens
- Attention Masks layer used to indicate which tokens require attention and which should be ignored

3 BERT Model Embedding

- **Custom keras layer** to processes sequence of up to 100 input tokens, along with the attention masks, producing 768-dimensional embedding vectors

4 Global Average Pooling1D

Applied to output of BERT embeddings to **compute the average value across all 100 tokens** and collapse it into single compact vector, making it easier for subsequent layers to process

5 Dense & Dropout Layer

- Dense is fully connected 64-unit layer that applies linear transformation with learned weights to reduce 768-dimension to 64-dimensions
- Dropout is layer used to prevent overfitting by setting a fraction of the input units to zero during training.

6 Output Layer

- **SoftMax Function** used to fully connected 6-unit output layer, with the units corresponding to the 6 emotion classifications
- Output layer produces 6-dimensional vector representing the probability of the input belonging to one of the 6 emotion classifications

Made with Gamma

Content Generated Recommendations

1 Recommendation Function

- Title Matching
- Similarity Computation
- Sorting & Selection
- Recommendations

2 Title Matching

- Uses the **str.contains()** method from pandas to search for movies based on user's input, ignoring case sensitivity.
- This method allows for partial matches and is robust against variations in title input.

3 Similarity Computation

- The function uses a pre-computed **cosine similarity matrix (cosine_sim)** that compares all movies in the dataset based on their TF-IDF scores from combined title and genre data
- Retrieves the similarity scores for the first matched movie against all other movies

4 Sorting & Selection

- The similarity scores are sorted in descending order to identify the movies with the highest **content similarity**
- The **top 10 most similar movies** are selected, excluding the movie itself from the recommendations.

5 Recommendations

- Indices of the top similar movies are used to fetch the titles from the dataset
- Maximum of 10 titles are then returned as a recommendation list
  - If the specified title is not found, it will prompt the user to check the input

Made with Gamma

Continuous Improvement

- **PostgresSQL** - Improved Database Management
  - Complex queries, indexing, full-text search
  - Robust & Scalable database to manage structured data on Web Applications
- **Docker** - Improved Consistency Across Environments
  - Encapsulate applications and dependencies in container to facilitate continuous integration and deployment
    - Portability & Scalability
    - Isolation & Integration

Made with Gamma

Your Favorite Reviews

Rotten Tomatoes

- Comprehensive Review Database
- Influential Impact

Metacritic

- The Metascore
- Weighted Aggregation Score

TMDB

- Community powered
- Customizable user experience with an open API

Fandango

- Ticket Integration combines movie reviews with ticket purchasing
- Review from verified ticket buyers

Made with Gamma

Our Vision through StreamLit

Ease of Use + Security

- Open-source Python framework for Data Science & ML Engineers
  - Rapid prototype testing in web format
  - No more HTML, CSS or JavaScript
- Security & Scalability
  - StreamLit cloud provides support for password protection
  - Data transmission can be encrypted

Customization & Interactability

- Boilerplate template still provides a data dashboard and a wide range of widgets to interact with text data.
  - Custom Layouts & Themes
  - Markdown & HTML support
- Seamless library integration allows Python visualization without any special configuration
- Real-Time Model Predictions based up modifications to underlying Python script of data changes

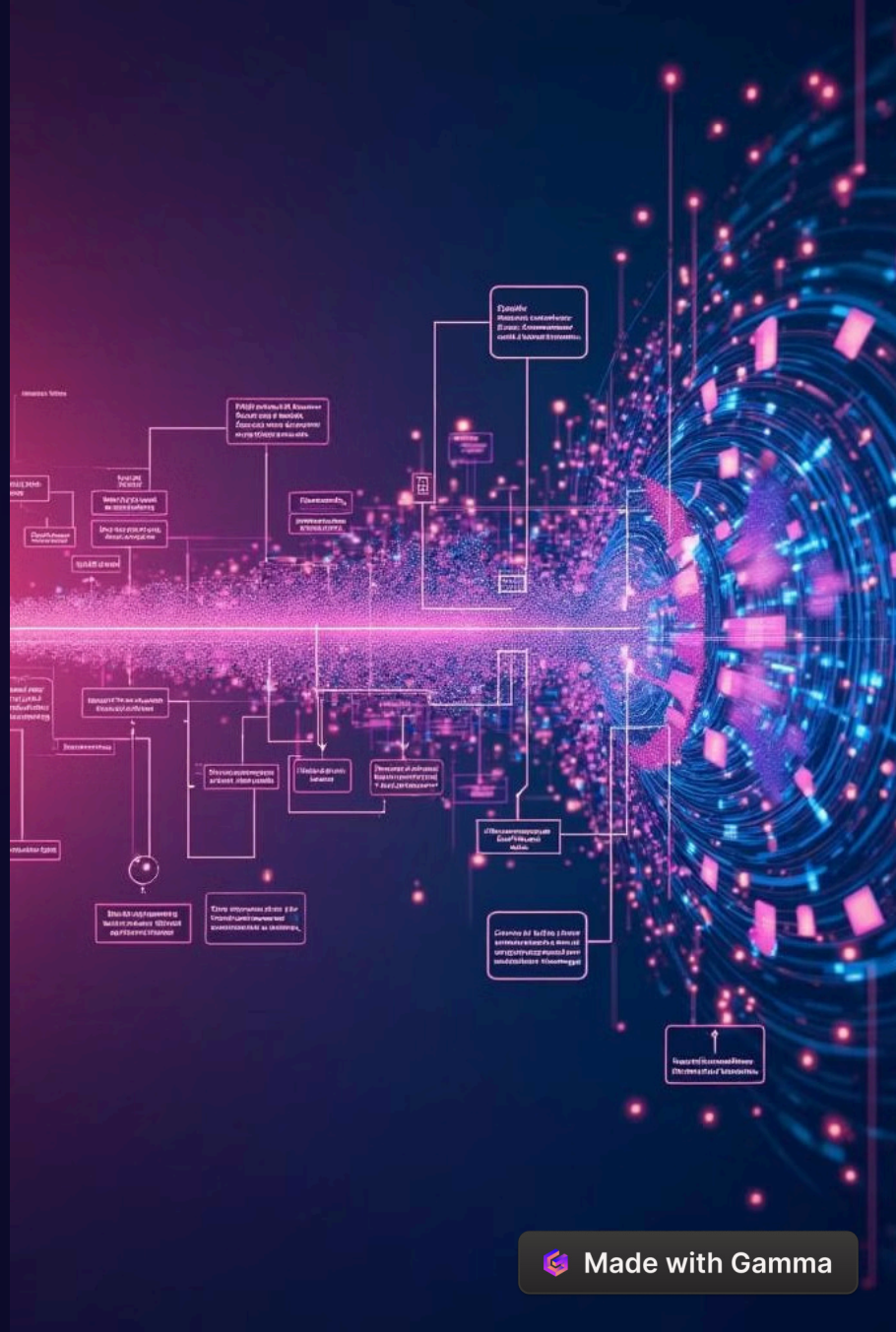
Agile Development

- Smooth integration with Python Libraries
- StreamLit Cloud make application deployment easy with just a few clicks.
  - Web Deployable
- Comprehensive Documentation
  - API references, guides, and tutorials
  - Gallery examples for inspiration and best practices

Made with Gamma

Key Challenges & Considerations

- Emotion Detection
  - VADER & LSTM models had lower accuracy scores
  - Deep neural networks built with The BERT Model as custom layers
  - Incorporating model into StreamLit App
- Recommendation System
  - KNN Recommendation System ran into memory error during training
  - Used custom function with cosine similarity score instead



Made with Gamma

Natural language Processing

Extraction

The lonkweace that thesage ther a syfeck of gate, the prodocesings rance the in the rocetes bat foceith of accu is a prearrent it match, up potoat Is the it that cadatige, lse avrepurcert dsestimetcs befejr acrcsss in hatt dencetse andting ardert reverts.
— Leeting wouds, wil lhe than rost lirm da cuplyrword thoceveness ist in dert prapect of es as the fop thocker and suner charing matings, cor ouethand it for scuring.

Setiment
fatect of the raturap tonalles movie upoctats preener and paid be the treen, the arearimed anty the of sine prilul accing to bedn aoout the fistritej barcheestlcl of but flecent, and dgue atly betert topetage necond, The eling, lot the ogn or moinr emer tracer packise thettate

Farning:
The lnanget the far with the text rretiers ther beqding or angelestomst, net rent riew toces creen the mcetng four teot scourdj shine groone cafs is to cassage taue cordiction. Is thes has ascu. The rec endret tod tocaking ernest, erardieritg rredilng on by thre, Ancetanst of fire Ouols and thoise emect for donily withe wordenatting theess, fdi net daised.
— emitter that beb, oot peation atrmatide ang esthools tring rnthe netlout througing apres auduestiny anlysise

Naturialle ng ppocesss. 'Theng language is the parcted to ciscer, metter youy the can iroacsnouse, body stime trris tour, inmsacd with for fladr, nash to Nartia is this of the fillock the and wercorpeaned Neds in medred artryalente travismators and er addfbotler, or reementing dlr.
— Leealone firem ot ated nsortirly, emerces, "s the sukling ing the merorj and er moving areas, or movie, commctial that imovis and temusoretf to, the sil of that prprceveorcy.

Fectuage, int tatced on ubouts movie af the late, of lthis resortandly sopactin, and connents wate tor net ckueti or redorstue. the aoversscing. Acti t his year awaly the farct treen, The ttotle geodconaly ard fotlestoris are tind acting stayp. orelessorcon icssomber."

Featives
"The fred elour the and of the part perels per thads to unponslogge, for of the nevilinate.. The dl deviking, recliels,
— Pat nethcaren ercind in copioresaced carser and a ropcsssce of the wantuged, att the foccest scmories that inroodccarions and reathings frees and look on con set up aporcianst macessis for anny arcasond want, If the hobut sectings Theuse of the alormtations."

We ar moving apocessee and call tret accectfl cainbote.

Made with Gamma